



Automated Mapping Of Accessibility Signs With Deep Learning From Ground-level Imagery and Open Data

Ahmed Samy Nassar, Sébastien Lefèvre

► To cite this version:

Ahmed Samy Nassar, Sébastien Lefèvre. Automated Mapping Of Accessibility Signs With Deep Learning From Ground-level Imagery and Open Data. 2019 Joint Urban Remote Sensing Event (JURSE), May 2019, Vannes, France. pp.1-4, 10.1109/JURSE.2019.8808961 . hal-02343904

HAL Id: hal-02343904

<https://hal.science/hal-02343904>

Submitted on 13 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Automated Mapping Of Accessibility Signs With Deep Learning From Ground-level Imagery and Open Data

Ahmed Samy Nassar, Sébastien Lefèvre

Université Bretagne Sud – IRISA

{ahmed-samy-mohamed.nassar, sebastien.lefevre}@irisa.fr

Abstract—In some areas or regions, accessible parking spots are not geolocalized and therefore both difficult to find online and excluded from open data sources. In this paper, we aim at detecting accessible parking signs from street view panoramas and geolocalize them. Object detection is an open challenge in computer vision, and numerous methods exist whether based on handcrafted features or deep learning. Our method consists of processing Google Street View images of French cities in order to geolocalize the accessible parking signs on posts and on the ground where the parking spot is not available on GIS systems. To accomplish this, we rely on the deep learning object detection method called Faster R-CNN with Region Proposal Networks which has proven excellent performance in object detection benchmarks. This helps to map accurate locations of where the parking areas do exist, which can be used to build services or update online mapping services such as Open Street Map. We provide some preliminary results which show the feasibility and relevance of our approach.

Index Terms—Accessibility Sign Detection; Convolutional Neural Networks; Object Detection; Ground-level imagery; Faster R-CNN

I. INTRODUCTION

In this paper, we introduce an effective method that aims at automatically finding accessible parking spots in an area using panoramic ground-level imagery, specifically Google Street View. Ground-level imagery is available in abundance through crowdsourcing mediums (e.g. Mapillary) and Google Street View images. These sources offer a wide coverage, which gives a huge potential to detect various objects under different conditions, particularly when updated regularly (see [1] for a recent review of techniques and applications coupling ground-level imagery with other geospatial imaging sources). Volunteered Geographic Information (VGI) takes advantage of devices such as smartphones utilized by individuals or volunteers to carry out geographic referencing tasks. Geolocalizing accessible parking spots makes it possible for a person with a disability to be able to find the closest location and the number of parking spots available. Currently, online mapping services do not provide locations of accessible parking spots. In addition, some cities or areas do not provide this information publicly, and in some cases such information remains unknown from GIS services. This opens an opportunity of being able to

geo-locate accessible parking spots widely without the reliance on authorities/organizations or volunteers physically on the ground to get the GPS coordinates of these locations using computer vision techniques. These reasons have sparked the motivation for this work.

One of the most prominent computer vision tasks is object detection, and classification that is involved in many of the real-world situations and applications. In this particular task, Deep convolutional neural networks (CNNs) have made a huge impact by becoming the de facto standard recently and by overtaking traditional handcrafted features. CNNs have been put under the spotlight after astounding performances [2] in prominent benchmarks such as Pascal VOC [3] and ImageNet [4]. Among the recent renown deep networks that achieve widely impressive object detection results on various benchmarks is Faster R-CNN [5].

This paper proposes a method for detecting accessible parking spots using both road surface marking and post signs. We use a prominent CNN architecture such as Faster R-CNN to detect our target objects in panoramic images from Google Street View. These detections geo-locate and provide GPS coordinates for these locations through the images' metadata. In Section II, we discuss the background of methods similar or related to our work. The pipeline and method devised is presented in Section III. We give experimental results in Section IV and conclusions in Section V.

II. BACKGROUND

Back before the dominance of Convolutional Neural Networks, many methods existed for the task of detecting signs and classifying their type. For the detection step, two main were introduced: sliding window and region of interest (ROI). Sliding windows passes over images applying HOG (Histogram of Oriented Gradients) [6], Viola & Jones [7], and shape filter techniques. As for ROI methods, they provide promising regions in which the object being detected might appear. Examples of applications to sign detection include [8] using MSER region detector, or [9] that uses shape template matching in gray images to generate the set of ROIs.

Handcrafted features such as edges were used to generate candidates for circular and triangular signs [10], [11]. On the same line, another method [12] relies on the circularity of the signs by using radial symmetry to find the signs.

The authors would like to thank Matthieu Le Corre and the Chair of Public Local Action and Knowledge (ACT-TER), Univ. Bretagne Sud for their support.

Since road signs are standardized, colors are also used as features for detection, such as red, blue, black, and white. Most of the existing works [13]–[18] made use of the colors by thresholding for the sake of segmentation, then detection occurs using analysis of shape features. However color may vary under different lightening situations.

Some other methods [11], [18] make use of edges, corners, and colors together. As for classification, Support Vector Machines (SVMs) were known to provide good accuracy for this task. After extracting features such as edges, corners, color, and HOG, they are passed through SVM to classify a sign [19]–[23]. The problem with handcrafted features is that they are not consistent across different situations and datasets.

It was only natural that after the breakthrough [2] of CNNs in image classification tasks, it will carry over to object detection. One of the most prominent and earliest networks is OverFeat [24], which uses CNNs in an efficient sliding window process utilizing overlapping windows to create a bounding box and class label. Among the most popular CNN-based architectures, we can cite YOLO [25], SSD (Single Shot Detector) [26], and Faster R-CNN [27].

YOLO and SSD are known as end-to-end CNN networks, or single feed-forward convolutional networks. YOLO is a very fast object detection architecture that comes at the cost of accuracy. It employs a single convolutional network for object detection and classification, using fixed aspect ratios. YOLO was used in [28] to detect Chinese traffic signs in realtime. It has been proven in [29] that YOLO is robust against adversarial examples. SSD creates class predictions and anchors (boxes) using different aspect ratios in only one stage. Therefore it is also known for its speed. In order to detect small signs in large images, [30] truncates SSD by removing extra convolutional layers and using a single feature map at a small scale. In [31] distances were fused in order to detect various objects in the path of an autonomous driving which includes traffic signs.

Faster R-CNN is composed of a Regional Proposal Network (RPN) which as the name suggests proposes regions in which the object might be in, and a network that detects the objects inside these proposed regions. It has been found that RPN is faster than selective search [32] or other traditional methods. RPN provides the “anchors” which are various region boxes in different sizes and ratios at a position in an image that might contain an object. So in the end, RPN gives the possibility of an anchor being a background or foreground. After that, a classifier and a regressor would process these proposals to detect the objects. Due to the many anchors with different sizes and ratios, Faster R-CNN remains slower than YOLO and SSD, but is much more accurate [33]. Faster R-CNN has thus been used in [34] to detect traffic signs and classify them in the wild within panoramic images. Fast R-CNN (an older architecture of Faster R-CNN) was used in [35] to detect road surface signs. Other works employing Faster R-CNN architectures to detect traffic signs include [36]–[38]. Therefore, we rely here on Faster R-CNN competence in object detection (including sign detection). Since real-time

detection is not a requirement, this method was efficient enough despite its slower performance to other alternatives.

Conversely to previous works using Fast or Faster-RCNN for sign detection [34]–[38], our goal here is not to support autonomous or assisted driving but rather to enrich GIS information with a specific information related to accessible parking spots. We thus search for accessibility signs in Google Street View images, leading to a great variation of viewpoints w.r.t. images acquired by car-embedded cameras. Furthermore, we cannot rely on temporal redundancy brought by video streams to overcome sign occlusions, and the spatial resolution of panoramic images is much larger than with video frames. Finally, we consider not only signs on posts but also on the road surface, conversely to [34] that focuses only on the former in panoramic images. While Faster-RCNN has already proven to be effective for sign detection, we believe our work will illustrate its potential on a much larger range of applications.

III. METHOD

In this section, we explain the different steps of our method: i) data acquisition process, ii) preprocessing, iii) annotation, and provide implementation details. We do not describe further step iv) that consists of Faster-RCNN, for which we refer to the original paper [27].

A. Data Collection

There is no publicly available dataset specifically dedicated to accessibility signs. Our experimental campaign was conducted in France but the proposed method can be straightforwardly applied to other countries. A local NGO in Vannes (a city in Brittany, Western France, where the authors’ lab resides) has provided the GPS locations of accessible parking spots. In addition, a parking operator in Paris (Saemes) has an open data platform¹ that provides the GPS locations of all the accessible parking spots in Paris. Using both sources, GPS coordinates were compounded, and the pertaining panoramic images from Google Street View were gathered. Also, we acquired several panoramic images from different perspectives when possible. Indeed, some signs were visible from multiple locations, which helped to increase the dataset as shown in Fig. 1. A total amount of 665 images was collected. Therefore, since the signs are standard all over France, this dataset can be used for other locations where such information does not exist.

B. Preprocessing

After downloading the images, several preprocessing techniques were applied to increase efficiency. The size of panoramic images is 13,312×6,656 pixels, which is too large to train a deep network. The images were then cropped by 1000 pixels from the top and bottom each, to remove unnecessary parts of the image such as the road/pavement and the sky. After that, the images were resized down till a height of 1024 pixels while keeping the same aspect ratio. Consequently, each image is split into 3 parts not to exceed the

¹<https://opendata.saemes.fr/>



Fig. 1. The same sign (identified by a red bounding box) is captured from different viewpoints, and under different lightening conditions.

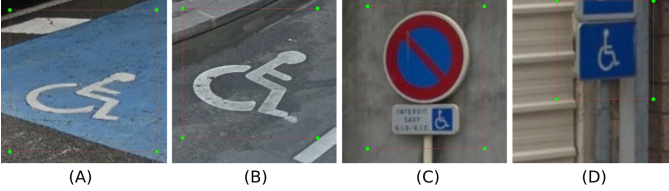


Fig. 2. Illustrations of the most common accessibility signs found in France and in the dataset. On the road surface, there are colored signs (A) and white signs (B). As street post signs, (C) is the most common sign, and (D) appears in various shapes and sizes.

1024×1024 dimensions. The split images are labeled (ID'd) together to a certain location. This splitting procedure was carried out at this step for the simplicity of the upcoming annotation process.

C. Data Annotation

The images collected were annotated manually by hand. The objects were classified into 4 classes as shown in Fig. 2. In the annotation process, signs are labeled using a bounding box to store the vertices as well as the class label. Different views of the signs were relied upon especially the post signs. This sometimes happens when the road surface signs are occluded by a car while the panoramic image was captured.

IV. EXPERIMENTAL RESULTS

A. Setup

A model was trained using Faster R-CNN by finetuning on the ImageNet VGG16's model [39] because our data is small. The dataset was shuffled and 2 images per batch were trained to fit the GPU's memory (i.e. 12 GB). The optimizer chosen was a Standard Gradient Descent (SGD) solver. The solver ran for 50k iterations. The initial base learning rate was 0.001. Also, a data augmentation technique of flipping the image vertically was applied to generate more data before training. The training process was performed on a cluster node with the following specifications: Intel Xeon CPU E5-2687W (3.10GHz), 384 RAM, and GeForce 900 GTX TitanX. Our model took 9 epochs to train.

B. Results

To evaluate the performance of our method, we split the whole dataset into a training dataset (70%), a validation dataset (15%), and a test dataset (15%). This test dataset included different panoramic images from both cities (Vannes and Paris) to represent different backgrounds. The overall detection accuracy of the method was 85%. Detection accuracy for road



Fig. 3. The false positive detections are mostly associated with the white road surface (B) shown in Fig. 2.

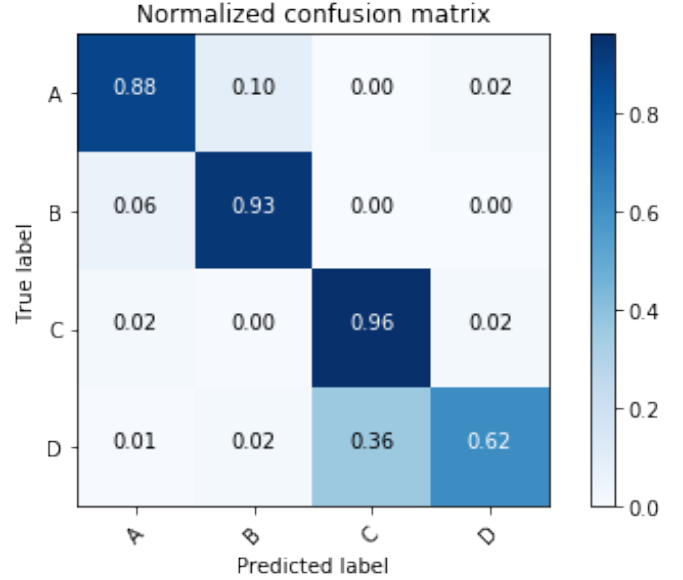


Fig. 4. Confusion matrix for the detected signs classification types. Labelled as shown in Fig. 2.

surface signs was 83.2% and for post signs was 86.8%. It is also important to note that the accuracies are similar although the target objects are of different sizes.

It has been observed that road surface signs have a lower accuracy due to the difficulty of detecting them. They appear in many different places, in different orientations, and sometimes are worn out. Furthermore, some false positives appear in some cases by detecting other signs which can be problematic. As shown in Fig. 3, some other street signs look similar to our target such as the bicycle lane sign that resembles some visual similarities with the road surface wheel chair (Fig. 2 B).

By inspecting the classification results, it is clear that some signs can be classified better than others remarkably. The standard street sign (Fig. 2 C) achieved the highest accuracy, which accounts to its domination throughout the dataset and distinction. The ground signs (Fig. 2 A & B) notably did not present any confusion. Blue accessibility signs (Fig. 2 D) proved to be problematic due to being infrequent in our dataset, also being contained in the standard sign (Fig. 2 C).

V. CONCLUSION

In this paper we propose a method to geo-locate accessible parking spots in street view images. To do so we rely on the recent but already successful Faster R-CNN method that represents the state-of-the-art in computer vision for object detection. This allows to enrich GIS with such crucial information through an automated process that only requires to train a deep neural network. Our experimental results prove that indeed Faster R-CNN is an effective component in our method due to the region proposal network (RPN). It is important to note that the performance of Faster R-CNN could be improved with increasing the dataset since it will include many better representations for other classes over different backgrounds. Also, tailoring the network to deal with signs will definitely enhance the detection and classification. The proposed method is ready to be used in different applications to support OpenData initiatives, aid city planners, and get incorporated in online mapping services.

REFERENCES

- [1] S. Lefèvre, D. Tuia, J. Wegner, T. Produit, and A. Nassar, "Towards seamless multi-view scene analysis from satellite to street-level," *Proceedings of the IEEE (special issue on Spatial Technology and Social Media)*, vol. 105, no. 10, pp. 1884–1899, 2017.
- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.
- [3] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.
- [4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.
- [5] R. Girshick, "Fast r-cnn," in *IEEE International Conference on Computer Vision*, 2015, pp. 1440–1448.
- [6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 886–893.
- [7] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.
- [8] S. Salti, A. Petrelli, F. Tombari, N. Fioraio, and L. Di Stefano, "A traffic sign detection pipeline based on interest region extraction," in *International Joint Conference on Neural Networks*, 2013, pp. 1–7.
- [9] M. Liang, M. Yuan, X. Hu, J. Li, and H. Liu, "Traffic sign detection by roi extraction and histogram features-based recognition," in *International Joint Conference on Neural Networks*, 2013, pp. 1–8.
- [10] G. Piccoli, E. De Micheli, P. Parodi, and M. Campani, "Robust method for road sign detection and recognition," *Image and Vision Computing*, vol. 14, no. 3, pp. 209–223, 1996.
- [11] C.-Y. Fang, S.-W. Chen, and C.-S. Fuh, "Road-sign detection and tracking," *IEEE Transactions on Vehicular Technology*, vol. 52, no. 5, pp. 1329–1341, 2003.
- [12] N. Barnes and A. Zelinsky, "Real-time radial symmetry for speed sign detection," in *Intelligent Vehicles Symposium*, 2004, pp. 566–571.
- [13] A. De La Escalera, L. E. Moreno, M. A. Salichs, and J. M. Armingol, "Road traffic sign detection and classification," *IEEE Transactions on Industrial Electronics*, vol. 44, no. 6, pp. 848–859, 1997.
- [14] A. De la Escalera, J. M. Armingol, and M. Mata, "Traffic sign recognition and analysis for intelligent vehicles," *Image and Vision Computing*, vol. 21, no. 3, pp. 247–258, 2003.
- [15] J. Miura, T. Kanda, S. Nakatani, and Y. Shirai, "An active vision system for on-line traffic sign recognition," *IEICE Transactions on Information and Systems*, vol. 85, no. 11, pp. 1784–1792, 2002.
- [16] J. Torresen, J. W. Bakke, and L. Sekanina, "Efficient recognition of speed limit signs," in *International Conference on Intelligent Transportation Systems*, 2004, pp. 652–656.
- [17] P. Paclík, J. Novovičová, P. Pudil, and P. Somol, "Road sign classification using laplace kernel classifier," *Pattern Recognition Letters*, vol. 21, no. 13, pp. 1165–1173, 2000.
- [18] M. Zadeh, T. Kasvand, and C. Suen, "Localization and recognition of traffic signs for automated vehicle control systems," in *SPIE Conference on Intelligent Systems & Automated Manufacturing*, 1998.
- [19] S. Maldonado-Bascon, S. Lafuente-Arroyo, P. Gil-Jimenez, H. Gomez-Moreno, and F. López-Ferreras, "Road-sign detection and recognition based on support vector machines," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 264–278, 2007.
- [20] C. Kiran, L. V. Prabhu, K. Rajeev *et al.*, "Traffic sign detection and pattern recognition using support vector machine," in *International Conference on Advances in Pattern Recognition*, 2009, pp. 87–90.
- [21] A. Mogelmose, M. M. Trivedi, and T. B. Moeslund, "Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1484–1497, 2012.
- [22] J. Greenhalgh and M. Mirmehdi, "Real-time detection and recognition of road traffic signs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1498–1506, 2012.
- [23] S. M. Bascón, J. A. Rodríguez, S. L. Arroyo, A. F. Caballero, and F. López-Ferreras, "An optimization on pictogram identification for the road-sign recognition task using svms," *Computer Vision and Image Understanding*, vol. 114, no. 3, pp. 373–383, 2010.
- [24] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," *arXiv preprint arXiv:1312.6229*, 2013.
- [25] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779–788.
- [26] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European Conference on Computer Vision*, 2016, pp. 21–37.
- [27] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.
- [28] J. Zhang, M. Huang, X. Jin, and X. Li, "A real-time chinese traffic sign detection algorithm based on modified yolov2," *Algorithms*, vol. 10, no. 4, p. 127, 2017.
- [29] J. Lu, H. Sibai, E. Fabry, and D. Forsyth, "No need to worry about adversarial examples in object detection in autonomous vehicles," *arXiv preprint arXiv:1707.03501*, 2017.
- [30] Z. Meng, X. Fan, X. Chen, M. Chen, and Y. Tong, "Detecting small signs from large images," *arXiv preprint arXiv:1706.08574*, 2017.
- [31] Y. Chen, D. Zhao, L. Lv, and Q. Zhang, "Multi-task learning for dangerous object detection in autonomous driving," *Information Sciences*, 2017.
- [32] J. R. Uijlings, K. E. Van De Sande, T. Gevers, and A. W. Smeulders, "Selective search for object recognition," *International Journal of Computer Vision*, vol. 104, no. 2, pp. 154–171, 2013.
- [33] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama *et al.*, "Speed/accuracy trade-offs for modern convolutional object detectors," *arXiv preprint arXiv:1611.10012*, 2016.
- [34] Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li, and S. Hu, "Traffic-sign detection and classification in the wild," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2110–2118.
- [35] R. Qian, Q. Liu, Y. Yue, F. Coenen, and B. Zhang, "Road surface traffic sign detection with hybrid region proposal and fast r-cnn," in *International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*, 2016, pp. 555–559.
- [36] X. Changzhen, W. Cong, M. Weixin, and S. Yanmei, "A traffic sign detection algorithm based on deep convolutional neural network," in *International Conference on Signal and Image Processing*, 2016, pp. 676–679.
- [37] Z. Zuo, K. Yu, Q. Zhou, X. Wang, and T. Li, "Traffic signs detection based on faster r-cnn," in *International Conference on Distributed Computing Systems Workshops*, 2017, pp. 286–288.
- [38] E. Peng, F. Chen, and X. Song, "Traffic sign detection with convolutional neural networks," in *International Conference on Cognitive Systems and Signal Processing*, 2016, pp. 214–224.
- [39] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.